

CLAIMS

What is claimed is:

1. A method of governing access to resources in a computing utility facility, comprising:

receiving a demand profile associated with an application that specifies a pattern

of resources from a pool of resources to be delivered with a class of service;

5 determining if the pool of resources has resources to be delivered to the

application at the specified class of service;

reducing the specified class of service to a lower class of service acceptable to the

application if the determination indicates the resource pool is unable to deliver the

resources at the specified class of service; and

10 admitting an application to the computing utility facility if resources delivered at

the reduced class of service are available from the pool of resources and acceptable to the

application.

2. The method of claim 1 further comprising,

assigning the resource from the pool of resources to the application in response to

15 admitting the application to the computing utility facility.

3. The method of claim 1 wherein the class of service is selected from multiple

classes of service including: a static class of service, a guaranteed time varying class of

service, a predictable best effort class of service and best effort class of service.

4. The method of claim 1 wherein determining if resources from the pool of

20 resources delivered at the specified class of service further comprises having the resources

available for delivery at a class of service at least as high as the class of service specified

by the application.

5. The method of claim 1 wherein the determination further comprises:

considering the demand requirements across multiple classes of service when

25 fulfilling the demand profile and class of service specified by the application.

6. The method of claim 1 wherein reducing the class of service further includes:

requesting that the application accept delivery of resources at a class of service lower than the class of service specified by the application according to the actual available resources in the pool of resources.

30 7. The method of claim 1 wherein the demand profile associated with an application further identifies the resources required from a pool of resources during one or more demand cycles.

8. The method of claim 1 wherein admitting the application further comprises:

35 unfolding the one or more demand cycles from the demand profile associated with the application into time slots requiring resources from the pool of resources at the requested class of service;

comparing the time slots requiring resources at the accepted class of service with a staging calendar of time slots representing the available resources in the pool of resources at the requested class of service; and

40 converting time slots from the staging calendar to a permanent calendar when the comparison indicates the time slots requiring resources from the demand profile are at the class of service requested and available for assignment.

9. The method of claim 8 wherein unfolding the one or more demand cycles includes a caveat time cycle and corresponding classes of service based upon events that occur 45 over a long-period of time and are selected from a set including special events, holidays, seasonal occurrences and emergencies.

10. The method of claim 9 wherein the caveat time cycle is based upon knowing when

at least one particular event is going to occur in the future.

11. The method of claim 8 wherein unfolding the one or more demand cycles
50 includes a demand cycle describing a demand for resources from a resource pool during weekdays and another demand cycle describing another demand for resources during weekends.
12. The method of claim 8 wherein converting time slots from the staging calendar to a permanent calendar comprises:

copying the time slots from the staging calendar to the permanent calendar; and
55 preallocating the requested resources from the pool of resources according to the permanent calendar schedule.

13. The method of claim 8 wherein converting time slots from the staging calendar to a permanent calendar comprises:

indicating the time slots in the staging calendar associated with the requested
60 resources are permanent and not for staging purposes; and
pre-allocating the requested resources from the pool of resources according to the permanent calendar schedule.

14. The method of claim 1 further comprising:

policing requests for resources from the admitted applications to determine if the
65 resources being requested in the accepted class of service are also within an acceptable range of demand.

15. The method of claim 14 wherein the policing further comprises:

intercepting a request for resources from an application admitted to access a pool
of resources;
70 determining if resource requested is within the accepted class of service and an acceptable range of demands based upon the demand profile of the application;

- indicating an application is not entitled to the request when the determination indicates the request is outside the acceptable range of demands and class of service; and indicating an application is entitled to the request when the determination
- 75 indicates an application is within the acceptable range of demands and class of service.
16. The method of claim 1 further comprising arbitrating the allocation of limited resources between two or more applications entitled to receive the limited resources.
17. The method of claim 16 wherein the arbitration further comprises:
- detecting a conflict in providing requested resources to two or more admitted
- 80 applications entitled to receive the limited resources;
- determining if at least one application can forego receiving the requested resources causing the conflict for a predetermined period of time according to a priority scheme;
- instructing the at least one application to forego receipt of the requested resources for a period of time in accordance with the determination;
- 85 allocating resources to the remaining admitted applications entitled to receive the requested resources in accordance with the priority scheme.
18. The method of claim 17 wherein the priority scheme includes selecting admitted applications to receive the requested resources on a first-come-first-serve basis.
19. The method of claim 17 wherein the priority scheme includes selecting admitted
- 90 applications to receive the requested resources according to economic and class of services factors.
20. The method of claim 19 wherein the priority scheme based upon economic and class of service factors further comprises:
- determining the potential economic penalties to the computing utility facility when
- 95 each of the two or more admitted applications does not receive the limited resources it is

entitled to receive;

ranking the two or more admitted applications entitled to receive the limited resources according to the potential economic penalties imposed upon the computing utility facility; and

100 selecting one or more admitted applications from the ranking to not receive the limited resources to minimize the potential economic cost to the computing utility facility.

20. An apparatus for governing access to resources in a computing utility facility, comprising:

a processor capable of executing instructions;

105 a memory containing instructions when executed cause the processor to receive a demand profile associated with an application that specifies a resource from a pool of resources to be delivered with a class of service, determine if the pool of resources has the resource to be delivered at the specified class of service to the application, reduce the specified class of service to a lower class of service acceptable to the application if the 110 determination indicates the resource pool is unable to deliver the resources at the specified class of service and admit an application to the computing utility facility if resources delivered at the reduced class of service area available from the pool of resources and acceptable to the application.

21. The apparatus of claim 20 further comprising instructions that,

115 assign the resource from the pool of resources to the application in response to admitting the application to the computing utility facility.

22. The apparatus of claim 20 wherein the class of service is selected from multiple classes of service including: a static class of service, a guaranteed time varying class of service, a predictable best effort class of service and best effort class of service.

120 23. The apparatus of claim 20 wherein the instructions that determine if a resource from the pool of resources delivered at the specified class of service further comprises instructions to determine if the resources available for delivery in a class of service are at least as high as the class of service specified by the application.

24. The apparatus of claim 20 wherein the instructions that determine further 125 comprise instructions that,

consider the demand requirements across multiple classes of service when fulfilling the demand profile and class of service specified by the application.

25. The apparatus of claim 20 wherein the instructions that reduce the class of service further includes instructions that,

130 request that the application accept delivery of resources at a class of service lower than the class of service specified by the application according to the actual available resources.

26. The apparatus of claim 20 wherein the demand profile associated with an application further identifies the resources required from a pool of resources during one or 135 more demand cycles.

27. The apparatus of claim 20 wherein instructions that admit the application further comprise instructions that,

unfold the one or more demand cycles from the demand profile associated with the application into time slots requiring resources from the pool of resources at the accepted 140 class of service, compare the time slots requiring resources at the accepted class of service with a staging calendar of time slots representing the available resources in the pool of resources at the requested class of service and convert the time slots from the staging calendar to a permanent calendar when the comparison indicates the time slots requiring

resources from the demand profile are at the class of service requested and available for
145 assignment.

28. The apparatus of claim 27 wherein the one or more demand cycles includes a
caveat time cycle and corresponding class of service based upon an event that occurs over
a long-period of time and is selected from a set including special events, holidays,
seasonal occurrences and emergencies.

150 29. The apparatus of claim 20 further comprising instructions that,
police requests for resources from the admitted applications to determine if the
resources being requested in the accepted class of service are also within an acceptable
range of demand.

30 The apparatus of claim 28 wherein the policing instructions further comprise
155 instructions that,

intercept a request for resources from an application admitted to access a pool of
resources, determine if resource requested can be delivered within the accepted class of
service and an acceptable range of demands based upon the demand profile of the
application, indicate an application is not entitled to the request when the determination
160 indicates the request is outside the acceptable range of demands and class of service and
indicated an application is entitled to the request when the determination indicates an
application is within the acceptable range of demands and class of service.

31. The apparatus of claim 20 further comprising instructions that arbitrate the
allocation of limited resources between two or more applications entitled to receive the
165 limited resources.

32. The apparatus of claim 31 wherein the arbitration instructions further comprise
instructions that,

- detect a conflict in providing requested resources to two or more admitted applications entitled to receive the limited resources, determine if at least one application 170 can forego receiving the requested resources causing the conflict for a predetermined period of time according to a priority scheme, instruct the at least one application to forego receipt of the requested resources for a period of time in accordance with the determination, and allocate resources to the remaining admitted applications entitled to receive the requested resources in accordance with the priority scheme.
- 175 33. The apparatus of claim 32 wherein the priority scheme includes instructions that select admitted applications to receive the requested resources on a first-come-first-serve basis.
34. The apparatus of claim 32 wherein the priority scheme includes instructions that select admitted applications to receive the requested resources according to economic and 180 class of services factors.
35. The apparatus of claim 34 wherein the economic and class of service factors further comprise instructions that,
- determine the potential economic penalties to the computing utility facility when each of the two or more admitted applications does not receive the limited resources it is 185 entitled to receive, rank the two or more admitted applications entitled to receive the limited resources according to the potential economic penalties imposed upon the computing utility facility and select one or more admitted applications from the ranking to not receive the limited resources to minimize the potential economic cost to the computing utility facility.
- 190 36. An apparatus for governing access to resources in a computing utility facility, comprising:

- means for receiving a demand profile associated with an application that specifies
a resource from a pool of resources to be delivered with a class of service;
means for determining if the pool of resources has the resource at the specified
195 class of service for delivery to the application;
means for reducing the specified class of service to a lower class of service
acceptable to the application if the determination indicates the resource pool is unable to
deliver the resources at the specified class of service; and
means for admitting an application to the computing utility facility if resources
200 from the pool of resources in the reduced class of service is acceptable to the application.